

**CSL** *COORDINATED SCIENCE LABORATORY*

## **UNDERSTANDING AND REPRESENTING NATURAL LANGUAGE MEANING**

DAVID L. WALTZ

APPROVED FOR PUBLIC RELEASE. DISTRIBUTION UNLIMITED.

**UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN**

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle)  Understanding and Representing Natural Language Meaning		5. TYPE OF REPORT & PERIOD COVERED  Annual Report 1/1/81 - 12/31/81
		6. PERFORMING ORG. REPORT NUMBER  T-114
7. AUTHOR(s)  David L. Waltz		8. CONTRACT OR GRANT NUMBER(s)  N00014-75-C-0612
9. PERFORMING ORGANIZATION NAME AND ADDRESS  Coordinated Science Laboratory University of Illinois at Urbana-Champaign Urbana, IL 61801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS  Office of Naval Research Arlington, VA		12. REPORT DATE  May 1982
		13. NUMBER OF PAGES  33
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)  UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  Artificial intelligence, Natural language processing, Cognitive modeling, Knowledge representation, Computer science		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  During this contract period, we: (a) introduced new representation schemes called "event shape diagrams" for events and actions; (b) developed mechanisms for "explanatory schema acquisition," a process which allows a program to learn new plans and scripts (i.e., action sequences) by reading articles or stories based on the plans or scripts; (c) completed work on computer programs which can understand complex noun phrases (such as "water pump pulley adjustment screw thread damage report summary"); (d) made substantial progress toward identifying and cataloging "cognitive universals" (processes and structures		

Annual Progress Report for Research in  
UNDERSTANDING AND REPRESENTING NATURAL LANGUAGE MEANING  
Contract No. N0014-75-C-0612

For the Period  
January 1, 1981 - December 31, 1981

Submitted to  
OFFICE OF NAVAL RESEARCH

Principal Investigator  
David L. Waltz

COORDINATED SCIENCE LABORATORY  
University of Illinois at Urbana-Champaign  
Urbana, IL 61801

May 1982



Progress Report for the Office of Naval Research

Contract N00014-75-C-0612

1.0 PERSONNEL

During the past reporting period the research group has had several changes. Dr. Gerald DeJong, who recieved his Ph.D. from Yale University in 1979, has joined the faculty and Dr. LaRaw Maran who is trained as a linguist has joined as a research associate. Our current graduate student research assistants are: Rick Dinitz, Jeff Gibbons, Paul O'Rorke, Jordan Pollack, and David Spoor.

2.0 SUMMARY OF RECENT PROGRESS

During this contract period, we: (a) introduced new representation schemes called "event shape diagrams" for events and actions; (b) developed mechanisms for "explanatory schema acquisition," a process which allows a program to learn new plans and scripts (i.e., action sequences) by reading articles or stories based on the plans or scripts; (c) completed work on computer programs which can understand complex noun phrases (such as "water pump pulley adjustment screw thread damage report summary"); (d) made substantial progress toward identifying and cataloging "cognitive universals" (processes and structures shared by all people, regardless of language or culture) through the comparison of English scene and event descriptions with similar descriptions in other languages; (e) completed our investigation of methods of evaluating natural language systems, and



applied the methods to our own PLANES system; (f) completed the JQL data base query language system, designed for easy connection to a natural language front end; and (g) completed HEXVIS, a parallel, message-passing system for high-level vision.

### 3.0 SUMMARY OF PROGRESS BY RESEARCH PROJECT

#### 3.1 Event Shape Diagrams - D. L. Waltz

We have made exciting breakthroughs in this area. For a long time, we have been both impressed by and dissatisfied with Schank's use of "primitives of conceptual dependency" [Schank, 1975] for representing the meaning of natural language. (a) Despite the demonstrated power of conceptual dependency (CD) the particular primitives used have changed over the years [Schank, et al., 1973; Schank and Riesbeck, 1981]. (b) The choice of primitives for use in definitions has apparently been an art, justified on the grounds of practical effectiveness, but never by theory. (c) The sets of primitives are incomplete [Wilks, 1975] -- they cannot represent certain verb meanings (e.g., divide, construct); they do not have distinguishable representations for many classes of verbs for example, (break, chip, crack, destroy, damage, and scratch would all be represented very similarly). More recently, Schank et al. seem to have had little interest in repairing defects of CD, but have concentrated instead on developing larger memory structures, in particular scripts [Schank and Ableson 1977] and MOPS [Schank 1979], each of which may contain a large number of CD structures.

Rieger's "common sense algorithm" (CSA) work [1975] attempted to enumerate causal relationships between states, actions, and tendencies (such as gravity) for use in representing the operation of physical mechanisms. While very promising in certain ways, CSA diagrams have never been well-integrated with CD, and have not been able to represent timing, quantitative state variable values, concurrency or hierarchical relationships in a satisfactory way. Furthermore, the construction of CSA

diagrams is still an art.

Adverbs (e.g., quickly, softly, hard, suddenly) have seldom been mentioned in AI papers on natural language understanding. When they have been mentioned [Cerccone, 1977], they have been viewed as difficult or impossible to deal with. We have developed mechanisms for dealing with adverbs [Waltz, 1981] and have recently improved upon these mechanisms. Many adverbs (including the ones above) can be represented very naturally in event shape diagrams, though manner adverbs (e.g., viciously, kindly, bluntly) still seem difficult.

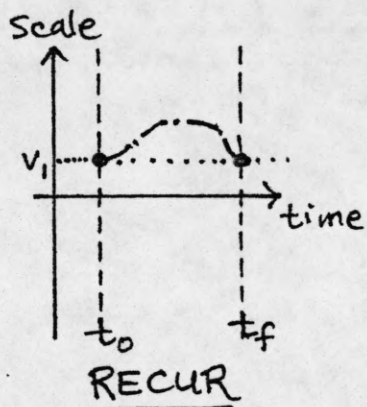
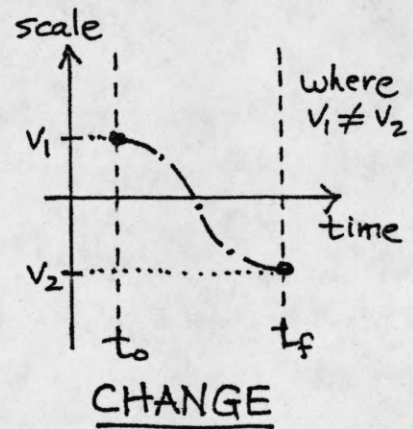
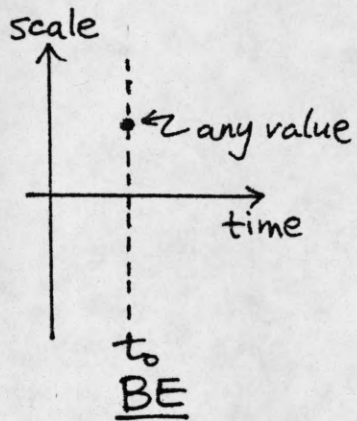
Recent work by Lehnert on summarizing narratives [Lehnert, et al., 1981], by Allen on a "temporal logic" [Allen, 1981], by Abelson on the relationships between events, actions, plans, and emotions [Abelson, 1981], and by Talmy on the relationships between grammar and perception [Talmy; 1978] have all played a part in the development of our "event shape diagrams," summarized below.

In their simplest forms, event shape diagrams have a time line, a scale, and values on the scale at one or more points. There are three basic event shape diagrams, illustrated in Figure 1.\*

---

\*While diagrams are shown here, data structures to represent these diagrams are very easy to program.





special case:  
 if  $\text{scale}(t) = v_1$   
 for all  $t$ ,  $t_0 \leq t \leq t_f$ ,  
 then the diagram represents  
PERSIST

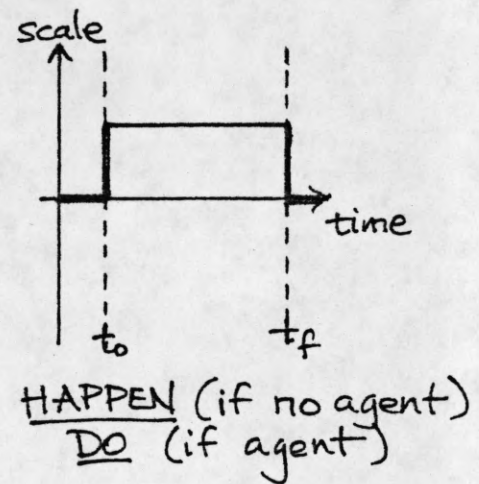


Figure 1. Basic event shape diagrams.

Diagrams can be used to represent concurrent processes, causation, and other temporal relations by aligning two or more diagrams, as illustrated in Figure 2. Figure 2 shows the representation for "eat." Note that four simple diagrams are aligned, and that each has different kinds of scales, and different event shapes. Causal relations also hold between the events described in each simple diagram. The names for the causal relations are adopted from Rieger's CSA work. The action Eating stops in this default case where "desire to eat" goes to zero. "Desire to eat" sums up in one measure coercion, habit, and other factors as well as hunger. Typical values for amounts of food, time required to eat, and so on are also associated with the diagram, to be used as default values.

More levels of detail can be added if needed. For instance, the action diagram can be expanded so that eating involves many recurrences of putting food in one's mouth, biting, chewing, and swallowing, and the diagram for the amount of food inside the agent can reflect a series of stepwise changes as each mouthful is ingested.

Many adverbial modifiers can be represented neatly: eat quickly shrinks the value of  $t_f - t_o$  with respect to typical values; eat a lot increases the values of  $q_o - q_f$  above typical values. Similarly "eat only half of one's meal," "eat very slowly," "eat one bite," etc. can be neatly represented.

The point of time from which events are viewed can also be clearly represented. Past tense (e.g. "we ate 3 hamburgers") puts "now" on the time line to the right of the action, while future tense puts "now" to the left of the action, and present progressive (e.g. "we are eating") puts "now" between  $t_o$  and  $t_f$ .

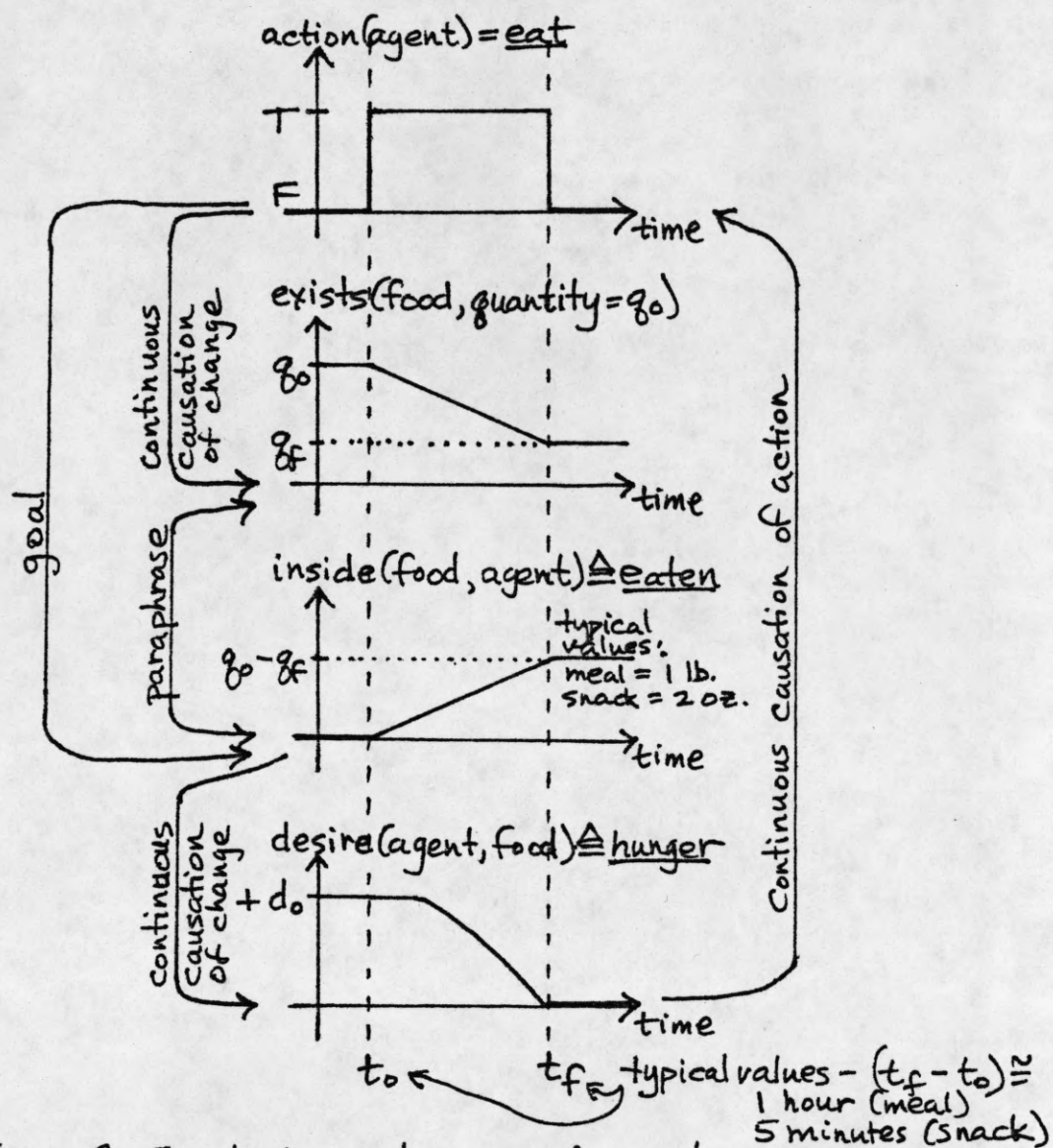


Figure 2. Event shape diagram for eat.



No AI systems have dealt with models of human belief, expectation, and attitude in any but the simplest situations [Cohen and Perrault, 1981]. While it is premature to make grandiose claims, the examples we have worked on so far have not presented great difficulties for our event shape diagram formulations. For example, Figure 3 shows the representation of the apparently rather hard sentence, "I was surprised that John ate so much." The structure in Figure 3 uses the portion of the preceding meaning for eat that is selected by the pattern eat + <quantity>. Be surprised has slots for an agent, an expected event or action, and an actual event or action, which must differ. In this case, we know that John actually ate more than he was expected to eat, so we can fill in his actual behavior in some detail. If the sentence were instead "I was surprised at how much John ate," it is also possible that John ate less than he was expected to, and the representation would simply show that the actual and expected amounts were different. The "interest arousal" scale shows up as part of the meaning of many verbs, such as like, enjoy, hate, pay attention to, desire, fear, and so on.

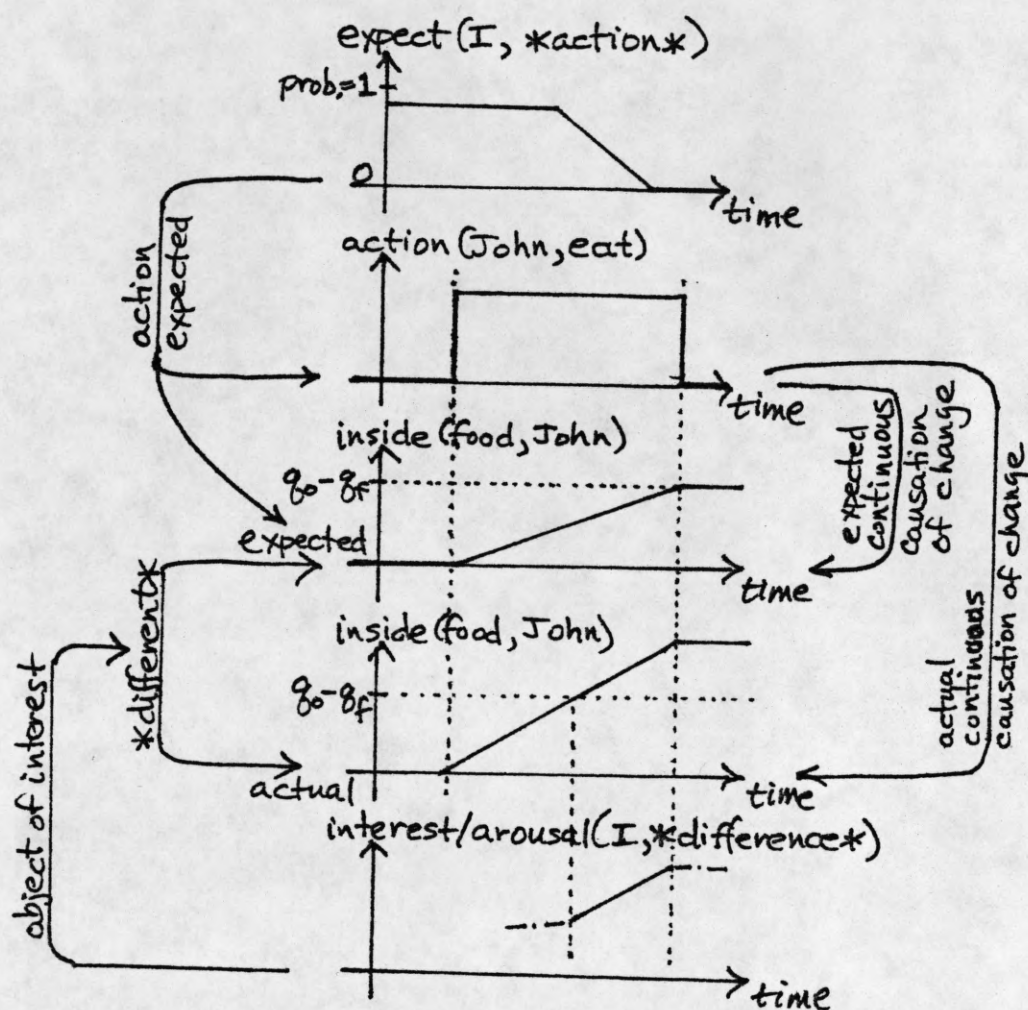


Figure 3. Representation of "I was surprised that John ate so much."

### 3.2 Explanatory Schema Acquisition - G. DeJong, P. O'Rourke

Over the past year we have been working on extending a natural language system, such as the one described in my [DeJong 1979], to learn how the world behaves directly from experiences. Natural language processing systems must have an immense amount of world knowledge in order to process input texts. We hope to have a computer system able to acquire some of this knowledge for itself.

The proposed learning method, called Explanatory Schema Acquisition, is a form of learning from direct experience. It will enable a system to acquire and refine new schemas. This will improve processing efficiency and extend the range of processing applicability.

The following example illustrates the kind of approach we are taking toward schema acquisition. In this example, from the domain of natural language understanding, our proposed system will learn the schema for kidnapping. We assume that the system does not yet have a schemas for kidnapping or extortion or any similar notion. As this is a knowledge based approach, the system does, however, possess a considerable quantity of background information about stealing, bargaining, the use of normal physical objects, and goals of people and institutions. Some of this knowledge, for example, knowledge about bargaining and stealing, must be in the form of schemas that the system already possesses. Example input story:

Paris police disclosed Tuesday that a man who identified himself Jean Maraneaux abducted the 12 year old daughter of wealthy Parisian businessman Michel Boullard late last week. Boullard received a telegram demanding that 1 million francs be left in a lobby waste basket of the crowded Pompidou Center in exchange for the girl. Asking that the police not intervene, Boullard arranged for the delivery of the money. His daughter



was found wandering blindfolded with her hands bound near his downtown office on Monday.

A KIDNAPPING schema, if the system had one, would contain information by which the system could judge the relative importance of and causal relationships between the story events. For example, a KIDNAPPING schema would enable the system to conclude that the transfer of 1 million francs is very important while the blindfolding is less so and that there is an important underlying connection between paying the money and the girl's freedom. If the system had such a schema, processing this story would be easy.

By assumption, the system does not have such a schema. However, in processing the story the system will be able to learn a general schema for processing kidnap stories from this one instance. Performance on succeeding kidnap stories will thus be greatly improved.

The general paradigm is to view understanding as the process of explaining input events. The explanations can then be used to generalize a single event into a new schema.

In processing this example without a KIDNAPPING schema the system can explain some events but not others. In particular, existing schemas cannot explain why Maraneaux might steal Boullard's daughter. While this is quite clearly an instance of taking something that belongs to someone else, there is no motivation for it. The daughter has no apparent value to Boullard, and a person, unlike money, cannot be used to acquire other valued goods. The system requires motivations for major volitional actions (such as a character invoking the STEAL schema). Therefore it is confused at this point in its processing.

The confusion is resolved by the next sentence. This input invokes the BARGAIN schema. The system understands the motivation for Maraneaux to initiate the bargaining event: he is trying to acquire money which it knows to be a possible goal of any human. Furthermore, this provides the motivation for the STEAL event. Maraneaux used the STEAL schema to satisfy the precondition of the BARGAIN schema of possessing the item to be traded.

Resolving the confusion causes the system to invoke its explanatory schemas acquisition procedure. This procedure does two things. First, it constructs a new schema composed of a STEAL event and a BARGAIN event where the STEAL is used to satisfy a precondition of the BARGAIN. Second, constraints on the slots for the new schema are derived from the knowledge in the systems STEAL and BARGAIN schemas and from the story as follows:

- 1) the slot filled by 1 million francs is generalized to be any amount of money.
- 2) the slot filled by Maraneaux (the kidnapper) is generalized to be any adult human.
- 3) the slot filled by the daughter is generalized to be anyone with close personal ties with (4).
- 4) the slot filled by Boullard is generalized to be any human who both has the amount of money to fill (1) and a person with close personal ties to fill (3).

Thus the system now has a schema that can be used to process a new story about a person stealing another person in order to trade him back for money. This is, of course, a first approximation to a schema for kidnapping.

We have identified four generalization techniques which, alone or in concert, seem to underly all of this form of learning. The four are: 1) schema composition, 2) secondary effect elevation, 3) schema alteration, and 4) non-volitional to volitional schema transformation. The above kidnapping example illustrates schema composition. One schema (STEAL) is

used to satisfy a precondition of another (BARGAIN). Secondary effect elevation involves using a known schema in such a way that what was previously a side-effect is now the main goal. For example, consider grand larceny arson. Normally the main goal of starting a fire is to destroy the object. Here, however, that is only a side-effect. The owner's collecting insurance money (which was only a side-effect) is now the main goal. Schema alteration involves changing an existing schema in some small, well defined way. For example, a schema describing how to turn in a screw can easily be altered to yield a schema for turning out a screw. Finally, non-volitional to volitional schema transformation makes a non-problem solving schema into a problem solving one by artificially supplying one of the known preconditions. For example, rain, normally an uncontrolled natural event, can be brought about by seeding the clouds. This research is very new and as yet the results are very tentative.



### 3.3 Understanding Metaphor - R. Dinitz

If (as some researchers have suggested) metaphor pervades the way we think, understand and explain our world, then an explanation of the processes we use (or could use effectively in a computer model) would be a significant contribution to our understanding of thought and memory.

Our approach to research in the subject of understanding meaning in metaphor can be broken down into four major divisions:

1. The Output Problem:

How could a program convince us that it had understood a metaphor?

2. The Learning Problem:

What happens when a program understands a metaphor that facilitates understanding of subsequent, related metaphors?

3. The Indexing Problem:

How would a program locate the information that might be useful in understanding metaphor?

4. The Transfer Problem:

How can relevant information be used for the understanding of a given metaphor?

1. In examining the output problem we must consider what kind of output we would require of a metaphor understander in order to convince us that:

- a) It was understanding the metaphors we fed it.
- b) It was learning from the processing it did.
- c) It represents a plausible model of human metaphor understanding.

The word "output" here refers not only to printed output proper, but also to any modifications the program makes to itself or its database. Thus the program itself, and its evolving structure, are considered to be part of the output -- especially in judging requirements 2 and 3.

2. Any work on the learning problem must account for the fact that in humans, processing of familiar metaphors takes less time and effort than the processing of new or unfamiliar metaphors. In other words, a computer program to understand metaphor, re-integrate that information into its network of concepts, and so be able to re-use that information to facilitate the processing of any future occurrences of other related metaphors. Some relevant questions are: How is the content of memory updated? What happens to old information in the memory?

3. The indexing problem is concerned with the process for finding the information needed to understand a metaphor. This is related to the process we call reminding, and to associative memory retrieval. Different facets of the problem are highlighted by questions such as: How do words refer to concepts? How are concepts related to other concepts, and to words which represent them? How do words and concepts remind us of other words and concepts? What kind of memory organization would facilitate a solution to the transfer problem?

4. In solving the transfer problem, we assume that information is stored with words and concepts in a network of schemata. The problem is how to use that information in the processing of a given input. Subproblems include limiting the extent of inferencing, choosing which properties are worth mapping to extract meaning from a metaphor in a given situation, mapping those properties between similar structures

(projection), and sensitivity to larger context.

Work in the area of metaphor during the past year includes investigations of attributive sentences, scalar mapping metaphors, and mapping of particle meaning.

Attributive sentences are those of the type:

X is Y  
 X is like Y  
 X is similar to Y  
 X can be compared to Y  
 Think of X as Y, etc.

In the past most researchers (primarily linguists and psychologists) have separated the analysis of attributive sentences into two cases -- those which are literal, and those which are metaphorical. This dissection seems unnatural, since the same sentence may convey both types of meaning depending on the context in which the sentence is spoken (or written). I set out to find a process that would understand attributive sentences -- handling literal and metaphorical cases in the same fashion. If all nouns are defined by a set of features (features are like predicates and values that apply to the noun), and each feature has a salience ranking (salience is a measure of how essential the feature is to the definition of the noun), then the meaning of an attributive sentence may be modeled by a mapping of features from Y to X. The high-salience features of Y are mapped over to the features of X. If they map onto high-salience features of X, then the sentence meaning is the predication of those features to X, and the sentence is understood to be literal. If they map onto low-salience features of X, then the effect is to boost their salience -- the meaning is again taken to be the predication of those features to X,



but the sentence is understood to have been used metaphorically.

Scalar map metaphors are really families between related metaphors, which all seem to flow from a single analogy of some range of experience and some other range of sensory-motor experience. For instance, the comparison of time with Euclidean 1-space (i.e., a time line) yields many common English metaphors: a point of time, a length of time, a segment of time, the continuum of time, etc. Imagine that we can take the first derivative of this mapping metaphor, and we find that we may talk about

$$\text{Rate} = (\text{Perceived Time})/(\text{Actual Elapsed Time})$$

in terms of

$$\text{Rate} = \text{Distance}/\text{Time}$$

For example: time marches on, the hours crawled by, time flies, time is running out, etc. Some significant questions are: which ranges of experience are most often used in scalar mappings? What kinds of transformations can be applied to these mappings in order to produce new mappings (e.g., the first derivative applied to "time is like 1-space" produces "movement of time is like movement in space")? Which classes of features are mapped in scalar mappings and which are left behind?

Particles in English are small words in English, used to modify the action described by a verb (e.g., "out" as in "blow out", "in" as in "turn in", and "through" as in "pick through"). Some verbs may be modified by a wide range of particles, each producing a different sense of the verb. Many researchers and system builders have tried to get around representation problems by designating each verb-particle pair as a

separate verb sense, but there may be a more elegant solution. Each particle when it stands alone, has a specific meaning which can be described by a set of notions associated with it. For example, "through" carries the notions of 'motion', that is a portion of 'space' traversed in an interval of 'time' along some 'path' situated in some 'medium' filled with 'material' in a 'pattern of distribution'. The meaning of "pick through" is not arbitrary, but rather largely determined by the interaction of the meaning of "pick" with that of "through". Similarly for "see through", "shine through", "think through", and "sit through", etc. If we can suitably determine the set of notions associated with other particles, we might develop an effective algorithm to achieve the mapping of these notions onto arbitrary verbs.

The three approaches above share some common threads. All three are dependent on some sort of mapping. Each mapping transfers features or notions from one domain to another and thereby derives some new meaning from the words under analysis. It is hoped that these mappings can be unified so they could be implemented by variations of a single algorithm. We would, of course, like to be able to argue that such an algorithm reasonably approximates processing in the human brain, and thus learn something about the nature of thought and people in the enterprise.

### 3.4 Computer Perception - D. Spoor

Computer vision systems in the past have tended to concentrate on the processing of images usually produced by television cameras. Such work has generally been called picture processing, image processing or image understanding. Computer visual perception goes beyond images of the world to encompass an understanding of the world. Such an understanding would involve both a computer model of the world it is perceiving, built from visual input, as well as general knowledge of the domain of perception. Image processing thus plays a role in computer perception, but is driven by the needs of the perceiving system.

The first step in building such a system is development of the visual input system. This system would start with raw visual input obtained from stereo television cameras. Conventional image processing techniques can then be applied to these images in order to extract the location and orientation of objects of interest. Central to this extraction will be the focusing of attention so that the entire scene need not be understood. Also of importance will be the synthesis of object location information from a variety of measures, both stereo and monocular. The output of such a system would be a model of what the current status of the observed scene, subject to the interest of the observing system. The representation system used in the model developed will depend upon the model used in the observing system.

The observing system will contain the current model of the domain situation as well as the knowledge for building and changing the model to conform to new visual input. The concept of time will be included in the model as both past structure and future possibilities are dependent on motion and hence time. Development of the structure for the model to be



used presents quite a challenge. High level models of the type described here for the most part do not exist in current vision systems. A primary goal of the model to be developed is that it should be useful to a natural language system. This means that the representation used must conform to the language system model at its most abstract level. In fact this constraint helps rather than hinders since language can suggest models of the physical world at the high levels where visual analysis provides little help. Because the model must be sufficient to represent the physical scene being observed the model must be complete in the sense that all configurations and objects observed must be representable. This should provide feedback to the natural language model designer and allow greater refinement of abstract representation of physical descriptions.

### 3.5 Inferential Understanding - J. Morgan

#### Synopsis

Our research is an exploratory study in the inferential interpretation of language input. The goal is to construct a theory that will, in conjunction with other systems, take input at user behavior and interpret that behavior by constructing a hypothesis of the user's goals and purposes as explanation for the behavior. This hypothesis then forms the basis of the computer's response. The exploratory study discussed here focuses primarily on the development of an inference system based on a logic of communicative actions.

#### Background

Present language-understanding and production systems, both those with practical orientation and those with theoretical ambitions, generally have been constructed to simulate intelligent behavior in a relatively narrow domain, with the result that generalization is either in principle impossible or depends crucially on significant theoretical progress on fundamental issues. Our long-range goal is to supplement such penetrating but narrow research by another strategy: by constructing a system whose initial design is to be as general and flexible as possible, in that it will attempt to emulate from the beginning what we believe to be involved in human language understanding (for full details of our view, see Morgan [in preparation]). One important feature of our approach is that humans (hence the system we hope to construct) understand and respond to the utterances of others not by being "driven" directly by the syntax or semantics of utterances, but by hypothesis they construct about the acts, goals, and purposes of the speaker in making the utterance. This, we

claim, is the kind of information that is stored in long-term memory, and this is the basis from which humans plan their actions in response to the utterances of others -- in particular, in responding to the needs and purposes of speakers as evidenced indirectly by their utterances. (See Morgan [1978] (TINLAP II) for some discussion.)

One clear illustration of the kind of thing we have in mind is the problem of indirect speech acts (see Searle [1975], Morgan [1978]), as exemplified by simple cases like

Do you have data on F4 accidents?

The response of a truly intelligent system would depend not only on the logical (yes/no) answer to the question, but on an evaluation of the user's needs in light of the nature of the data. If the data is small, then a reasonable response might be (depending on what has gone before in the interchange) to present all the data as a response; otherwise an intelligent system might inquire further as to what part of the data the user might like to see, and by what medium and method it should be presented. While it is easy to see that an ad hoc subsystem could be constructed to display just these properties for a limited range of input problems, we wish to explore methods for providing general, principled strategies that would entail such responses in a systematic way. This involves, we claim, the capacity to make inferences about the goals, purposes, etc. -- in short, the plan structure -- behind the user's utterances.



## Methods

Our present research focuses on this aspect of understanding. It assumes (counter-factually) that present systems of meaning/knowledge representation like predicate calculus or "conceptual dependency" are sufficient to represent the "literal meaning" of utterances, and is directed mainly at the problem of constructing a theory of how humans infer plans/purposes/etc. behind speech acts.

We hypothesize (as do most researchers on this question) that humans in linguistic interchange construct models of each other's models of the world, and update these models on the basis of each utterance as it occurs in the discourse. A system with this ability must have at least:

- some amount of general "knowledge of the world"
- knowledge of typical properties of human beings
- principles of common-sense reasoning

It seems clear, as well, that an adequate inference system must have some kind of "inference driver" -- in other words, that the inference process is somehow goal-directed, rather than proceeding blindly and deductively until all possible inferences have been made.

The goal of the present research, then, is to begin exploring how such a theory can be fleshed out and articulated, by attempting to instantiate our initial hypotheses in LISP programs, with the exception that "debugging" such programs will demonstrate where the gaps in our knowledge and the flaws in our theories lie.

### 3.6 Cognitive Universals - L. R. Maran

The problems of research to which we (i.e., Waltz and myself) have addressed our major attention concern the apparently perplexing relationship between language and cognition. We are convinced that an initial clarification of this relationship may well lead to significant developments within AI efforts to represent knowledge. I bring to this effort a background in anthropology, linguistics, and psychology, knowledge of a number of southeast Asian languages (especially Jinghpaw, the language of a hunter-gatherer people of northern Burma), and a general perspective not much represented within AI.

It is widely acknowledged that developing a systematic means of knowledge representation is central to the whole AI enterprise; however, to say how one can determine what an appropriate approach to knowledge description and representation has remained an enigmatic question.

First of all, we have the meaning of the real-world on the one hand, and the representational mechanisms of natural language on the other; the latter must, by virtue of its representational role, abbreviate, classify and generalize (i.e., organize) the details of the incredibly complex and comprehensive real-world meaning correlates. The basic elements of language (whether sentential, morphemic or phonemic) and those of the real-world meaning correlates (whether conceptual or perceptual) cannot be in a one-one correspondence relationship. Moreover, each natural language embodies a specific system of representing meaning that is not necessarily shared in any uniform sense by others. Each natural language system organizes aspects of reality in a way that sets it apart, and because of this its approach to cognition very likely reflects the organizational bias inherent in that particular approach, and if so, it will follow that the

world-view it projects is neither essentially complete nor impartial.

The following discussions of specific problems indicate both the accomplishments of the last year as well as the work we are currently engaged in research.

### 3.6.1 The Encoding Of Spatial Meaning -

The prepositional phrase (PP) is the constituent in the surface syntactic configuration of a spoken sentence (S) wherein spatial-locative information is generally specified. In the example S1 in English, the parenthesized component constitutes the PP.

(S1) A bird is sitting "on the peak of the roof of my house."

We have been primarily interested in the real-world information encoded into the nominal expression (or noun phrase) within the PP; we have made some progress in sorting out cognitive universal categories of spatial meaning in this type of phrase.

An important lesson we have learned from Jinghpaw (JP), is the fact that language level representation of spatial meaning is very different in different natural language systems. For instance, locative meaning may be expressed by syntactical mechanisms or by processes which are nonsyntactical. The following examples show a basic contrast in spatio-physical meaning concerning whole-part relationships -- e.g., a house, the roof of the house, the peak on the roof of the house.

(S2) That house surely has a nice peak on its roof.

(S3) This peak used to be on the roof of the house over there.

(S4) We saw a grand old house with a fantastic peak on the roof



in Duluth last summer.

(S5) Remember that grand peak on the roof of the Athenium? Well, it disappeared when the building was remodeled recently.

The relationship between the peak and the house-roof in S2 and S4 implies a physically adhering or continuing relationship; S3 and S5 on the other hand, characterize a peak-to-house-roof relationship that has been physically altered. The first type of physical whole-part relationships is COMPOSITIVE, whereas the second is PARTITIVE. The important feature to note in English is the fact that the specification of these two categories of spatial meaning are not syntacticized. The choice of appropriate verbal forms -- has, used to be, disappeared -- and prepositions -- on, with -- and the critical use of the speaker's (or performative) context wherein perceptual schemata play crucial roles; "that house, this peak, over there," are verbal schemata that depend on visual-gesticulative codes.

In Jinghpaw, the compositive meaning of physical whole-part relationship employs a lexically-oriented compound-like approach, e.g.

(S6) nta magaw machyun -- "house-roof-cresting" ;  
house roof cresting

whereas, for the partitive meaning, the nouns are obligatorily modified to install a genitival or possessive relationship, e.g.

(S7) nta-a magaw-a machyun -- "The cresting of the roof of the  
house roof cresting house."  
POSS POSS

The fundamental differences underlying the English-JP encoding strategies illustrate for us two basic facts about language and cognition; first, different natural languages are in effect different verbal encoding

strategies; and second, in order to account for natural language understanding as such it is necessary to proceed from a foundation of comparative approaches. The realization that certain aspects of cognition, such as the compositive/partitive meanings of physical whole-part relationships, may be represented by syntactical grammar as in JP, and yet by vastly different mechanisms in others, as in English, is extremely important. Grammatical processes representing cognitive phenomena are not uniform from language to language; therefore, notions of grammatical universals are seriously inadequate in representing the meaning of natural language. Despite the fact that AI researchers have avoided the trap of looking to syntax for ideas on cognitive organization, AI researchers have not looked beyond English meaning encoding.

### 3.6.2 Exteriorizing-Interiorizing Perceptions -

JP also accentuates, within the Noun Phrase in the PP, its systematic accommodation of the real world by transparently setting up a sequence of nominals to represent an exteriorized perception of the location of a given event, and a second sequence of nominals to represent an interiorized view of the location. To each sequence of these nominals JP attaches an appropriate set of prepositions of spatial meaning.

(S8) shi bawlung-hpe nta-nhku yupgawk-kata-de  
 he ball Obj. house-inside bedroom-inside-toward  
                                    $N_1$            $P_1$            $N_2$            $P_2$

htawng dat ai -- "He kicked the ball toward the bedroom of the  
 kick release house."

The nominal sequence  $N_1$  specifies an object name, "house," as if that object has been viewed from some distance -- hence the exteriorizing sense of perception -- followed by the preposition  $P_1$  which states that the inside of the house is involved in providing a location for the event. This sequence is followed by  $N_2$  which names the interior unit/compartment of  $N_1$  together with  $P_2$  which specifies either an exact location (at, on, in) or a direction (to, toward). JP provides us with a real-world model that is quite precise vis-a-vis, the English approach.

The point to note here is that by means of a string of noncommutable nominals and prepositions JP represents perceptual information that would, in English, be left to pragmatics or to shared knowledge. In other words, from the standpoint of perceptual specificity in information encoding, some languages are relatively more thorough than the others. In order to develop a viable theory of the perceptual basis of meaning understanding we will need to draw heavily from such languages. We continue our research in this area with strong commitment.

### 3.6.3 Final Remark -

Our current research on spatio-temporal understanding has followed, and benefitted from, previous work done by members of our group, i.e. Boggess (1978: Computational Interpretation of English Spatial Prepositions; Report 'T-75), Waltz and Boggess (1979: Visual Analog



Representations for Natural Language Understanding; WP-20), Waltz (1979: Relating Images, Concepts, and Words; WP-23), and Waltz (1980: Generating and Understanding Scene Descriptions; WP-24).

Our approach to natural language data is now substantially expanded by the inclusion of non-English languages, and we have begun to have a clearer picture of the problems underlying the language-cognition relationship. We expect to be able to suggest some general models of cognition and natural language understanding during the current year where the added insights from our comparative approach play crucial roles.

### 3.7 Parallel Network Natural Language Processing - J. Pollack

The decomposition of language into syntax, semantics and pragmatics has never been a clean one, but has enabled researchers to address the representation and processing of each form of linguistic knowledge without regard for the others. Although the human brain is organized in a massively parallel fashion, the serial computer metaphor with a central processing unit and fixed-address memory has been the dominating influence in models of intelligence and language performance. (It was long thought that language could be processed serially through a syntax "box", a semantics "box", then a pragmatics "box".) It seems clear, now, that language is processed in parallel, with all three (and maybe more) sources of knowledge integrated in the decision processes [Schank and Birnbaum, 1980; Marslen-Wilson, 1980]. However, there is a paradox in knowledge systems due to serial implementation: Although the systems should get faster when given more knowledge, they get slower! Perhaps, then, the key to human-like performance of natural language processing is parallel organization.

One promising and quite general type of parallel organization is a relational network coupled with the twin iterative processes of spreading activation and lateral inhibition. Variations of this "activation/inhibition network" organization have been used to model human perception of letters and words [Rumelhart and McClelland, 1981] and to explain aspects of human memory [Collins and Quillian, 1972; Fahlman, 1979] and memory priming [Collins and Loftus, 1975; Ortony, 197x]; in one case an activation/inhibition network has even been used as part of a hypothetical organization of the mind [Minsky, 1980]. Whether the nodes in

the network are used to represent perceptual hypotheses (as in Rumelhart and McClelland), conceptual schemata (as in Collins and Quillian, etc.), or active agents (as in Minsky), activation and inhibition seem to be adequate and appropriate mechanisms for finding the best hypothesis, most fitting schema, or strongest agent, in a parallel fashion, and without central control.

The research being proposed here will attempt to establish that an activation/inhibition network organization is indeed useful in processing natural language. The nodes of the network in this project will be hypothesized partial meaning-structures as well as predictions (or expectations) for meaning-structures.

The main problem in using activation networks lies in determining an appropriate structure for the network; i.e. in the instantiation and connection of nodes. Minsky has been criticized for not addressing this problem; other researchers such as Ortony and Rumelhart and McClelland assume an a priori network structure. This particular research project will instantiate nodes for an activation/inhibition network in a breadth-first manner.



#### 4.0 CHANGES IN COMPUTING ENVIRONMENT

Progress has been made in several areas involving the computer software and hardware in use. With conversion to a VAX 11/780 computer now underway, an excellent programming environment has been developed for this machine. Conversion to the new machine should be quick and involve a minimum of disruption.

With the arrival of the first VAX 11/780 at CSL the process of conversion from our old DEC-10 was begun. Franz Lisp rather than MACLISP is supported on the VAX 11/780 under UNIX/V32. It was discovered that Franz Lisp is indeed close to MACLISP and anything that ran under MACLISP should run under Franz Lisp. Soon after conversion was started it became evident that the EMACS editor developed at CMU by James Gosling was superior to the standard VI editor. EMACS was obtained from CMU and installed on the VAX 11/780. EMACS provides a LISP syntax oriented editor as well as a screen management system that allows all programming to be done from inside EMACS. Soon after EMACS was up the need for greater terminal display capability was apparent. A search of currently available terminals was made and the Ann Arbor Ambassador was selected as best fitting our needs. Able to display 48 lines the Ann Arbor Ambassador allows several windows of reasonable size to be displayed at one time. It also has a meta-function key greatly spreading editing.

Since CMU has transferred much of their software to their VAX 11/780s, similar functions to those installed in MACLISP on the DEC System-10 were obtained, along with expanded capabilities. Thus essentially the same editor and debugger available in MACLISP are available in Franz Lisp. A similar capability file management package was also installed, as well as

extra system functions making Franz Lisp compatible with CMULisp.

The entire VAX 11/780 system appears to offer great promise. The vast main memory, and speedy disk input/output greatly speeds program development. EMACS along with the CMU software relieve the programmer of many tedious tasks allowing him to concentrate more on problem solving.



## REFERENCES

- Abelson, R. 1981. Personal communication.
- Allen, J. F. 1981. "An interval-based representation of temporal knowledge," Proc. IJCAI-81, Vancouver, B.C. Canada, Aug. 1981, pp. 221-226.
- Black, M. 1962. "Metaphor," in Models and Metaphor, Cornell University Press, Ithaca, NY, 1962.
- Bogges, L. C. 1979. "Computational interpretation of English spatial prepositions," Tech. Rept. T-75, Coordinated Science Laboratory, University of Illinois, February.
- Carbonell, J. G. 19xx. "Metaphor: An inescapable phenomenon in natural language comprehension," Technical Report CMU-CS-81-115, Carnegie-Mellon University, Department of Computer Science.
- Cercone, N. 1977. "A note on representing adjectives and adverbs," Proc. IJCAI-77, MIT, Cambridge, MA, Aug. 1977, pp. 139-40.
- Collins, A. and E. F. Loftus. 1975. "A spreading activation theory of semantic processing," Psychological Review, vol. 82, pp. 408-428.
- Collins, A. and M. R. Quillian. 1972. "Experiments on semantic memory and language comprehension," in L. W. Gregg (Ed.), Cognition in Learning and Memory, Wiley, New York.
- DeJong, G. 1979. "Skimming stories in real time: An experiment in integrated understanding," Research Report 158, Yale Computer Science Department, New Haven, CT.
- Fahlman, S. E. 1979. NETL: A System for Representing and Using Real-World Knowledge, MIT Press, Cambridge, MA.
- Lakoff, G. and M. Johnson. 1980. Metaphors We Live By, University of Chicago Press, Chicago.
- Lehnert, W. G., J. B. Black and B. J. Reiser. 1981. "Summarizing narratives," IJCAI-81, Vancouver, B.C., Canada, Aug. 1981, pp. 184-189.
- Marslen-Wilson, W. and L. K. Tyler. 1980. "The temporal structure of spoken language understanding," Cognition, vol. 8.
- McClelland, J. L. and D. E. Rumelhart. 1980. "An interactive activation model of the effect of context in perception," Tech. Repts. 91 and 95, Center for Human Information Processing, USCD.
- Minsky, M. 1980. "K-lines: A theory of memory," Cognitive Science, vol. 4, pp. 117-133.



Morgan, J. 1978. "Toward a Rational Model of Discourse Comprehension," TINLAP II, Urbana, IL, July, 1978, pp.109-114.

Ortony, A. 1975. "Why metaphors are necessary and not just nice," Educational Theory, Winter.

Ortony, A. 1979. Metaphor and Thought, Cambridge University Press.

Ortony, A. 197x. "SAPIENS: Spreading activation processing of information enclosed in associative network structures," unpublished.

Perrault, C. R. and Cohen, P. R. 1981. "It's for your own good: A note on inaccurate reference," in Joshi, Sag, and Webber (Eds.), Elements of Discourse Understanding, Cambridge University Press, pp. 217-230.

Rieger, C. 1975. "The commonsense algorithm as a basis for computer models of human memory, inference, belief and contextual language comprehension," in R. Schank and B. Nash-Webber (Eds.), Theoretical Issues in Natural Language Processing, ACL, Arlington, VA, pp. 180-195.

Schank, R. C. 1973. "Identification of conceptualizations underlying natural language," in R. C. Schank and K. M. Colby (Eds.), Computer Models of Thought and Language, Wott and Freeman, San Francisco, pp. 187-247.

Schank, R. C. 1975. "The primitive ACTs of conceptual dependency," in R. Schank and B. Nash-Webber (Eds.), Theoretical Issues in Natural Language Processing, ACL, Arlington, VA.

Schank, R. C. 1979. "Reminding and memory organization: An introduction to MOPs," Research Report 170, Computer Science Dept., Yale University, December.

Schank, R. C. and R. Abelson. 1977. Scripts, Plans, Goals, and Understanding. Lawrence Erlbaum Associates, Hillsdale, NJ.

Schank, R. C. and L. Birnbaum. 1980. "Memory, meaning, and syntax," Reseach Rept. 189, Yale Computer Science Dept., November.

Schank, R. C., N. Goldman, C. J. Rieger and C. Riesbeck. 1973. "Margie: Memory, analysis, response generation, and inference on English," 3rd Intl. Joint Conf. on Artificial Intelligence, August, Stanford Univ., pp. 255-61.

Schank, R. C. and C. Riesbeck. 1981. Inside Computer Understanding. Lawrence Erlbaum Associates, Hillsdale, NJ.

Sutherland, I. E. and C. A. Mead. 1977. "Microelectronics and computer science," Scientific American, vol. 237, no. 9, pp. 210-228, September.

Talmy, L. 1978. "The relation of grammar to cognition - A synopsis," in D. L. Waltz (Ed.), TINLAP-2: Theoretical Issues in Natural

Language Processing-2, Association for Computing Machinery and The Association for Computational Linguistics, pp. 14-24.

Talmy, L. 19xx. "Rubber-sheet cognition in language" Proceedings of the 13th Regional Meeting of the Chicago Linguistic Society.

Waltz, D. L. 1979. "Relating images, concepts, and words," Proc. of the NSF Workshop on the Representation of 3-D Objects, University of Pennsylvania, Philadelphia.

Waltz, D. L. 1980. "Generating and understanding scene descriptions," in Joshi, Sag and Webber (Eds.), Elements of Discourse Understanding, Cambridge University Press; also Working Paper 24, Coordinated Science Laboratory, University of Illinois, Urbana, February.

Waltz, D. L. 1981. "Toward a detailed model of processing for language describing the physical world," Proc. IJCAI-81, Vancouver, B.C., Canada, Aug. 1981, pp. 1-6.

Waltz, D. L. and L. C. Boggess. 1979. "Visual analog representation of natural language understanding," Proc. 6th Intl. Joint Conf. on Artificial Intelligence, vol. 2, Tokyo, Japan, August, pp. 926-934.

Whorf, B. L. 1956. Language, Thought and Reality, MIT Press, Cambridge, MA.

Wilks, Y. A. 1975. "Primitives and words," in R. Schank and B. Nash-Webber (Eds.), Theoretical Issues in Natural Language Processing, ACL, Arlington, VA, pp. 38-41.